A Charge Control Relation for Bipolar Transistors

By H. K. GUMMEL

(Manuscript received August 1, 1969)

We give a relation which links emitter and collector junction voltages, V_{eb} and V_{cb}, collector current I_c, and the total charge Q_b of carriers that enter through the base terminal (electrons in a pnp *transistor):*

$$I_c = \text{const}\,\frac{e^{(qV_{eb}/kT)} - e^{(qV_{cb}/kT)}}{Q_b}.$$

This relation is valid for high injection conditions, subject only to minor restrictions. The significance for device modeling is discussed.

I. INTRODUCTION

A basic concept of charge control theory is that the controlled current (collector current) equals controlling charge (base charge) divided by a transit time.[1] This paper presents an additional relation which links base charge and collector current with junction voltages. The validity of this relation is subject only to minor restrictions. When these are met the relation holds even under high-level injection conditions.

Section II presents a derivation of the new charge control relation. Equation (15) states the principal result. The discussion in Section III points out the significance of this relation for bipolar transistor models.

II. DERIVATION

Consider a one-dimensional transistor of pnp polarity. The hole current density is given by

$$j_p = q\mu Ep - qDp' \tag{1}$$

where the symbols have their customary meaning. We shall assume that diffusivity D and mobility μ are related through the Einstein relation:

Approximation (a):

$$D = \frac{kT}{q}\mu.$$

Approximation (b_1): It is assumed that electric fields are low enough for avalanche multiplication of carriers to be negligible.

Approximation (b_2): The velocity-field relation is idealized by the field dependent mobility expression:

$$\mu = \frac{\mu_o}{1 + \frac{\mu_o |E|}{v_s}}$$

where $\mu_0 \equiv qD_0/kT$ is the low-field mobility, considered for convenience independent of doping, and where v_s is the scattering limited velocity. Approximation (b_1) places an upper limit on allowable bias. It is known (see for example Ref. 2) that D is underestimated at high fields by approximations (a) and (b_2) and that approximation (b_2) yields too gradual a transition from low field velocities to the high field saturated velocity.[3,4] Nevertheless, approximations (a) and (b) afford significant simplifications in the treatment to follow and are retained for that reason. To the extent that our final result, equation (15), is affected by them, it must be considered approximate. The errors depend on bias and doping profile; they are not expected to exceed a few percent for typical situations. The error due to approximation (b_2) overemphasizes velocity saturation effects and may be alleviated by choice of values of v_s larger than the final saturation value in high field regions. In high-field regions the current is carried predominately as drift current, with a carrier concentration that is nearly constant in such regions, so that errors in D are of minor consequence.

Next we define a quantity $a(x)$ which is the ratio of the hole current density at position x to the current density j_c leaving the collector terminal

$$a(x) = \frac{j_p(x)}{j_c}. \tag{2}$$

For direct current conditions, considered here, a approaches unity at the collector and is $1/\alpha = (1 + \beta)/\beta$ at the emitter. For large common emitter current gain β, a differs negligibly from unity.

We use

$$E = -\frac{kT}{q}\psi' \tag{3}$$

where ψ is the electrostatic potential in units of the Boltzman voltage kT/q, and consider equation (1) as a differential equation for $p(x)$. Its solution, when p is specified at a point x_1, is

$$p(x) = p(x_1)e^{\psi(x_1)-\psi(x)} - \frac{j_c}{qD_o}\int_{x_1}^{x} a(t)e^{\psi(t)-\psi(x)}\,dt - \frac{j_c}{qv_s}\int_{x_1}^{x} a(t)\mid\psi(t)'\mid e^{\psi(t)-\psi(x)}\,dt. \tag{4}$$

Equation (4) is valid for any pair of points x_1 and x. Denote by x_E and x_C the outside edges of emitter and collector transition regions, and use equation (4) with $x_1 = x_E$ and $x = x_C$. Multiplication of equation (4) by $e^{\psi(x)}$ and use* of

$$p(x) = n_i e^{\varphi_p(x)-\psi(x)} \tag{5}$$

$$n(x) = n_i e^{\psi(x)-\varphi_n(x)} \tag{6}$$

where φ_p and φ_n are hole and electron quasi-fermi levels in units of the Boltzman voltage, yield

$$j_c = \frac{qD_o n_i^2(e^{\varphi_p(x_E)} - e^{\varphi_p(x_C)})}{\int_{x_E}^{x_C} a(t)n_i e^{\psi(t)}\,dt + \frac{n_i D_o}{v_s}\int_{x_E}^{x_C} a(t)\mid\psi'(t)\mid e^{\psi(t)}\,dt}. \tag{7}$$

We shall now show that the second term in the denominator is negligible. The integrals in the denominator obtain the largest contribution from the region near x_m where $\psi(x)$ attains its maximum value ψ_m. If in the second integral we replace $a(t)$ by its value a_m at x_m, and if we neglect $e^{\psi(x_E)}$ and $e^{\psi(x_C)}$ in comparison with e^{ψ_m}—all very reasonable assumptions—we obtain for the second integral in the denominator

Approximation (c):

$$\frac{D_o n_i}{v_s}\int_{x_E}^{x_C} a(t)\mid\psi'(t)\mid e^{\psi(t)}\,dt \approx \frac{2a_m D_o n_i}{v_s}e^{\psi_m}.$$

For an assessment of the relative magnitude of the terms in the denominator of equation (7), consider that in a region of width w the potential $\psi(x)$ does not differ markedly from ψ_m; such region is conventionally called the "base" of the transistor. Consider high current gain, that is, $a \approx a_m \approx 1$. Then the value of the first integral is $wn_i e^{\psi_m}$, compared with $(2D_0/v_s)n_i e^{\psi_m}$ for the second. The quantity $2D_0/v_s$ has units of length and is ≈ 200 Å for silicon. This length is small compared

* Equation (6) is defined for later reference.

to base widths of today's most advanced transistors and hence we will neglect the second term in the rest of this paper. Conceivably, future transistors may have narrow enough bases that the term will have to be kept.

Approximation (d):

$$\frac{n_i D_o}{v_s} \int_{x_E}^{x_C} a(t) \mid \psi'(t) \mid e^{\psi(t)}\, dt \ll \int_{x_E}^{x_C} a(t) n_i e^{\psi(t)}\, dt.$$

If in equation (7) we were to let $v_s \to \infty$, that is, considered the carrier velocity to be strictly proportional to the electric field, then approximation (d) would be implemented automatically. Note, however, that in making approximation (d) we do not imply $v_s = \infty$, nor do we neglect essential consequences of the finiteness of v_s. A low value of v_s manifests itself in substantial base widening at high currents, that is, in influencing $\psi(t)$ in the remaining (first) term in the denominator of equation (7). In view of the idealized textbook treatments in which the minority carrier concentration at the base side of the collector depletion region is set equal to zero, rather than to a finite value, the following statement of equation (7) may be of interest: For low injection (that is, for currents sufficiently low that the base width is independent of current) the effect of the finiteness of the scattering limited velocity on the dc collector current is equivalent to a base widening of $2D_0/v_s$.

We now make the approximation:

Approximation (e):

> The value of the electron quasi-fermi level in the base is constant.

A gradient in the electron quasi-fermi level in the region where electrons are majority carriers would cause appreciable electron current to flow; for transistors of reasonable current gain, such currents are negligible. Thus, approximation (e) is very reasonable. We denote this value of the electron quasi-fermi level by φ_{nb} and divide numerator and denominator of the right side of equation (7) by exp (φ_{nb}). We define the emitter-base and collector-base junction voltages by

$$V_{eb} = \frac{kT}{q} [\varphi_p(x_E) - \varphi_{nb}] \tag{8}$$

$$V_{cb} = \frac{kT}{q} [\varphi_p(x_C) - \varphi_{nb}]. \tag{9}$$

These voltages differ from terminal voltages by ohmic drops, primarily lateral ohmic drops in the base region. The first integral in the denominator of equation (7), after it is divided by exp (φ_{nb}), contains very nearly the total area density of electrons.

Approximation (f):

$$\int_{x_E}^{x_C} n_i a(t)(e^{\psi(t)-\varphi_{nb}})\,dt = \int_{x_E}^{x_C} a(t)n(t)\,dt.$$

The integrands outside the base region differ, since there the quasi-fermi level is position dependent and does not equal φ_{nb}, but the contribution to the integral outside of the base region is negligible. By defining an average value $\langle a\rangle_{av}$ of a,

$$\langle a\rangle_{av} = \frac{\int_{x_E}^{x_C} a(t)n(t)\,dt}{\int_{x_E}^{x_C} n(t)\,dt}. \tag{10}$$

We may write expression (f) as

$$n_i \int_{x_E}^{x_C} a(t)e^{\psi(t)-\varphi_{nb}}\,dt = -\frac{\langle a\rangle_{av}}{q}\,q_b \tag{11}$$

where q_b is the total charge, per unit area, of those mobile carriers associated with the base terminal, that is, electrons in a pnp transistor. Equation (7) with approximations (d) and (f) may be written

$$j_c = -\frac{(q^2 D_o n_i^2/\langle a\rangle_{av})[e^{(qV_{eb}/kT)} - e^{(qV_{cb}/kT)}]}{q_b}. \tag{12}$$

We now change from current and charge densities to current and charge. We chose the sign of the collector current according to the convention that an electric current entering the device is positive:

$$I_c = -j_c A \tag{13}$$

$$Q_b = q_b A \tag{14}$$

where A is the device area. Note that the sign of Q_b is such that an electric current flowing into the base tends to increase Q_b. This is the proper sign for charge control theory. Equation (12) can now be written

$$I_c = C\frac{e^{(qV_{eb}/kT)} - e^{(qV_{cb}/kT)}}{Q_b} \tag{15}$$

with

$$C = \frac{(qn_iA)^2 D_e}{\langle a \rangle_{av}}. \tag{16}$$

Equation (15) is the principal result of this paper. Note that Q_b depends on bias, and that the form of the bias dependence is governed by the doping profile. However, the relation among the quantities I_c, V_{eb}, V_{cb}, and Q_b in equation (15) is independent of the details of the doping profile, provided that assumptions (a) through (f) are valid.

III. DISCUSSION

In spite of the simple appearance of equation (15)—indeed because of it—it provides a powerful tool for transistor modeling. It may be written in the form

$$I_c = -a_{21}[e^{(qV_{eb}/kT)} - 1] + a_{22}[e^{(qV_{eb}/kT)} - 1] \tag{17}$$

with

$$a_{22} = -a_{21} = \frac{C}{Q_b} \tag{18}$$

which is the form of one of the Ebers–Moll equations.[5] But whereas in the Ebers–Moll equation the coefficients a_{21} and a_{11} are constant, they depend according to equation (18) on bias through the base charge Q_b (C depends on bias only through $\langle a \rangle_{av}$ which for $\beta \gg 1$ is nearly unity and varies little with bias). It is this bias dependence of Q_b which contains high-injection effects. Thus, use of equation (15) holds promise for transistor modeling of improved accuracy. The major bias dependence of the collector current is through the exponentials in the numerator of equation (15). These are "ideal" exponentials (unity emission coefficients) and involve no approximations. The actual modeling is now done on Q_b in the denominator. A bipolar transistor model using this approach will be presented in a later paper.

REFERENCES

1. Beaufoy, R., and Sparkes, J. J., "The Junction Transistor as a Charge-Controlled Device," ATE J., *13,* No. 4 (October 1957), pp. 310–324.
2. Persky, G., and Bartelink, D. J., "High Field Energy Distribution and Diffusion Coefficient for Heavy Holes in p-Germanium," Physics Letters, *28A,* No. 11 (March 10, 1969), pp. 749–750.
3. Prior, A. C., "Field-Dependents of Carrier Mobility In Silicon and Germanium," J. Phys. Chem. Solids, *12,* No. 2 (January 1960), p. 175–183.
4. Seidel, T. E. and Scharfetter, D. L., "Dependence of Hole Velocity Upon Electric Field and Hole Density for p-Type Silicon," J. Phys. Chem. Solids, *28,* No. 12 (December 1967), pp. 2563–2574.
5. Ebers, J. J. and Moll, J. L., "Large Signal Behavior of Junction Transistors," Proc. IRE, *42,* No. 12 (December 1954), pp. 1761–1772.